DOCUMENT RESUME

ED 311 074 TM 013 911

AUTHOR Byrne, Barbara M.
TITLE Teacher Burnout: Validating the Maslach Burnout Inventory across Four Educational Levels.
PUB DATE Aug 89
NOTE 36p.; Paper presented at the Annual Meeting of the American Psychological Association (97th, New Orleans, LA, August 11, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Elementary Secondary Education; Factor Analysis; *Factor Structure; Foreign Countries; Higher Education; Instructional Program Divisions; Intermediate Grades; *Professors; Psychometrics; *Secondary School Teachers; *Teacher Burnout; *Test Validity
IDENTIFIERS Between Group Differences; Canada; Confirmatory Factor Analysis; Exploratory Factor Analysis; *Maslach Burnout Inventory

ABSTRACT

The factorial validity of the Maslach Burnout Inventory (MBI) was studied for 98 elementary school, 163 intermediate school, 162 secondary school, and 218 university teachers in Ottawa (Canada). The equivalence of factorial measurements and structure was tested across the groups. Exploratory factor analyses were used to examine the initial factor structure, and confirmatory factor analyses tested the hypothesized three-factor structure and its invariance across groups. In spite of the findings of a few aberrant items specific to each group, the MBI demonstrated remarkably sound psychometric properties with intermediate and secondary school teachers, less sound properties with university teachers, and poor psychometric properties with elementary school teachers. Tests of invariance across intermediate and secondary teachers revealed the equivalence of all item measurements for both groups. The results have important implications for substantive studies focusing on multigroup comparisons across teaching levels. Five tables summarize study findings. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED311074

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BARBARA BYRNE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Teacher Burnout: Validating the Maslach Burnout Inventory Across Four Educational Levels

Barbara M. Byrne

University of Ottawa

Paper presented at the American Psychological Association Annual Meeting, New Orleans, August 1989

The author wishes to thank both the teachers at the Carleton Board of Education and the professors at the Univerity of Ottawa for giving freely of their time in order to participate in this study.

TM013911

BEST COPY AVAILABLE

Abstract

The purposes of the study were: (a) to test for the factorial validity of the Maslach Burnout Inventory (MBI), for 641 teachers at the elementary (n=98), intermediate (n=163), secondary (n=162), and university (n=218) levels, and (b) to test for the equivalence of factorial measurements and structure across groups. Exploratory factor analyses were used to examine the initial factor structure, and confirmatory factor analyses to test the hypothesized 3-factor structure, and its invariance across groups. Despite findings of a few aberrant items, specific to each group, the MBI demonstrated remarkably sound psychometric properties with intermediate and secondary teachers, less so with university professors, and poorly with elementary teachers. Tests of invariance, across intermediate and secondary teachers, revealed the equivalence of all item measurements, albeit only one factor covariance. The study has important implications for substantive studies focusing on multigroup comparisons across teaching levels.

Teacher Burnout: Validating the Maslach Burnout Inventory Across Four Educational Levels

The purposes of the present study were twofold: (a) to test for the factorial validity of the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981a, 1986), separately, for teachers at the elementary, intermediate, secondary and university levels, and (b) to test for the equivalence of factorial measurements and structure across these four levels of educators.

Burnout, a term first introduced by Freudenberger (1974), denotes the inability to function effectively in one's job as a consequence of prolonged and extensive job-related stress. It is considered the final step in a progression of unsuccessful attempts to cope with negative stress conditions (see Selye, 1956). The syndrome is most strongly linked to those who work in the human service professions, including social workers, teachers, nurses, lawyers, physicians, police officers, and others who engage in a high degree of contact with people in need of aid (see Maslach, 1982; Perlman & Hartman, 1982 for a review). Furthermore, the syndrome appears to be most critical for those who work in institutional settings (Caplan & Jones, 1975; Ianni & Reuss-Ianni, 1983; Tosi & Tosi, 1970).

The seminal research of Maslach and colleagues was the first of an empirical nature to investigate the phenomenon of burnout (for an historical summary, see Maslach, 1981a; Maslach

& Jackson, 1984). Their findings were consistent in supporting a multidimensional construct comprising three related, yet independent components: (a) emotional exhaustion -- feelings of fatigue that develop as one's emotional energies become drained, (b) depersonalization -- the development of negative and uncaring attitudes toward others, and (c) reduced personal accomplishment -- a deterioration of self-competence, and dissatisfaction with one's achievements.

These three elements of burnout have been empirically validated for teachers at the elementary, intermediate, and secondary levels of the educational system (Beck & Gargiulo, 1983; Gold, 1984; Iwanicki & Schwab, 1981; Jackson, Schwab, & Schuler, 1986; Schwab & Iwanicki, 1982a, 1982b). Teachers exhibit signs of emotional exhaustion when they perceive themselves as unable to give of themselves to students, as they did earlier in their careers. Depersonalization is evidenced when teachers develop negative, cynical and sometimes callous attitudes towards students, parents and colleagues. Finally, teachers reflect feelings of reduced personal accomplishment when they perceive themselves as ineffective in helping students to learn, and in fulfilling other school responsibilities. Overall, teachers who fall victim to burnout are likely to be less sympathetic toward students, have a lower tolerance for classroom disruption, be less apt to prepare adequately for class, and feel less committed and dedicated to

their work (Farber & Miller, 1981).

The development of the MBI was based on samples of workers from a wide range of human service organizations, including nurses, teachers, police officers, physicians, social workers, psychologists, psychiatrists, and lawyers (Maslach & Jackson, 1981a, 1986). The instrument comprises 22 items designed to measure both the frequency and intensity of burnout as represented by three components -- emotional exhaustion (EE), depersonalization (DP), and personal accomplishment (PA). Findings related to teachers, however, have revealed a high correlation between the frequency and intensity dimensions of the construct (see e.g., Gold, 1984; Iwanicki & Schwab, 1982); it has therefore been recommended that future research bearing on teacher burnout consider only the frequency dimension.

As a consequence, in large part, of Schwab and Iwanicki's (1982a, 1982b; Iwanicki & Schwab, 1981) work related to burnout among teachers, Maslach and Jackson (1986), in collaboration with Schwab, developed the Educators' Survey (MBI Form Ed), a version of the MBI specifically designed for use with teachers. The MBI Form Ed measures the same three factors of burnout as the original version of the MBI; the only difference between the two versions lies in the modified wording of certain items to make them more appropriate to a teacher's work environment. Specifically, the generic term "recipient", used in the MBI to refer to clients, has been replaced by the term "students".

Exploratory factor analyses of the MBI have yielded three well-defined burnout factors representing EE, DP, and reduced PA for a variety of helping professions in general (Firth, McIntee, McKeown, & Britton, 1985; Maslach & Jackson, 1981b), and for teachers in particular (Beck & Gargiulo, 1983; Belcastro, Gold, & Hays, 1983; Gold, 1984; Iwanicki & Schwab, 1981). Each of the subscales measuring these factors, based on the frequency dimension, has demonstrated strong evidence of (a) internal consistency reliability with alpha coefficients ranging from .76 to .90 (mean α = .78) (Beck & Gargiulo, 1981; Belcastro et al., 1983; Iwanicki & Schwab, 1981; Maslach & Jackson, 1981b), (b) test-retest reliability, with coefficients based on a 2 to 4-week interval ranging from .60 to .82 (mean r = .74) (Maslach & Jackson, 1981b), (c) convergent validity with external criteria including personal experience (observations), dimensions of job experience, and personal outcomes (Maslach & Jackson, 1981b, 1986), and (d) discriminant validity as evidenced by low and nonsignificant correlations between MBI scores, and job satisfaction and social desirability (Jackson et al., 1986; Maslach & Jackson, 1981b, 1986).

Taken together, these findings provide strong support for the MBI as a potentially reliable and valid measure of teacher burnout. However, as noted by Maslach and Jackson (1986), more construct validity research is needed in order to more fully establish the psychometric soundness of the instrument. In this

regard, at least three limitations in previous work are evident. First, although claims of a validated 3-factor structure of burnout have been reported, the MBI itself, has not been directly validated for teachers at either the secondary or university levels. Second, factorial validity of the MBI has been examined using only an exploratory approach; several deficiencies associated with these procedures are now well documented (see e.g., Alwin & Jackson, 1980; Fornell, 1983; Long, 1983; Marsh & Hocevar, 1985); confirmatory factor analysis has been shown to provide a more powerful test of factorial validity. Finally, while Gold (1984) has claimed evidence of factorial invariance across similar samples of teachers, such claims are considered specious from a statistical perspective. Although the literature suggests that the factorial structure of the MBI may vary across and within professional occupational groups (see Belcastro et al., 1983; Firth et al., 1985; Iwanicki & Schwab, 1981), the assumption of equivalent factorial validity has not been directly tested.

The present study addressed these issues by (a) testing the factorial validity of the MBI for teachers at the elementary, intermediate, secondary and university levels, (b) using a confirmatory factor analytic approach to validity procedures based on the analysis of covariance structures, and (c) testing for the factorial invariance of the MBI across teachers at each of four levels of the educational system.

Method

Sample and Procedure

Participants in the study were 641 teachers from 6 elementary (n=98), 6 intermediate (n=163), and 4 secondary (n=162) schools, and one university (n=218) in Ottawa, Canada. By necessity, data collection procedures differed for the non-university and university samples. For the former, the questionnaires were administered and collected in person by the investigator during a special meeting arranged by the principal of each school; participation was voluntary and restricted to only full-time teachers. For the university sample, full-time professors were randomly selected from a master list comprising all faculties. Copies of the questionnaire, together with instructions and a return envelope, were mailed to each faculty member.

Instrumentation

The MBI (Form Ed; Maslach, Jackson, & Schwab, 1986) was used to measure burnout for teachers at the elementary, intermediate, secondary, and university levels. The 22-item instrument is structured on a 7-point fully anchored scale ranging from 0 "feeling has never been experienced", to 6 "feeling is experienced daily". The EE, DP, and PA subscales comprise nine, five, and eight items, respectively.

Analysis of the Data

Data analyses were based on a three-stage process. First,

exploratory factor analyses (EFAs) were conducted separately for elementary, intermediate, secondary, and university samples. Second, confirmatory factor analytic (CFA) procedures using the LISREL VI computer program (Joreskog & Sorbom, 1985) were conducted to test, more rigorously, the hypothesized 3-factor structure underlying the MBI; each teacher sample was tested separately. Finally, the factorial invariance of the MBI was tested, statistically, across elementary, intermediate, secondary, and postsecondary teachers. These analyses included tests for the equivalence of (a) item and subscale measurements related to the number of factors, pattern of factor loadings, and scaling units, and (b) theoretical structure of the burnout construct.

In covariance structure analysis, determination of the extent to which a proposed model fits the observed data should entail multiple criteria (Fornell, 1983; Joreskog, 1979; Long, 1983; Marsh & Hocevar, 1985). Assessment of fit in the present study was based on (a) the chi-square (χ^2) likelihood ratio, (b) the χ^2 /degrees of freedom ratio; a value <1.50 for a sample size of 1000 indicates a reasonable fit to the data (B. Muthen, personal communication, January, 1987), (c) the Bentler and Bonett (1980) normed index (BBI), (d) the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973)[1], and (e) T-values (parameter estimates relative to their standard errors of estimate), normalized residual values and modification indices, all

provided by the LISREL program.

Results

Exploratory Factor Analyses

The EFAs, in general, supported previous work in demonstrating a well-defined factor structure; target loadings for each teacher group are presented in Table 1. Target loadings refer to factor loadings on the factor that the item was designed to measure; these loadings should be substantial and preferably >.30. Non-target loadings refer to all other loadings, and are expected to be small (<.10).

Insert Table 1 about here

Three aberrant items require some elaboration here; interestingly, these items appear to be group-specific. For example, in Table 1 we see that for elementary teachers, Items 10 and 11 loaded minimally on the expected DP factor (.17, .19, respectively); these items loaded substantially on the EE factor (.43, .64, respectively). For all other teacher groups, Items 10 and 11 loaded as expected, although Item 11 cross-loaded onto the EE factor (.34) for the university group.

In a similar fashion for intermediate teachers, Item 16 loaded minimally (.11) on the expected EE factor, but strongly (.74) on the DP factor. For the other teacher samples, this item loaded adequately on EE, albeit other group-specific

peculiarities are worthy of note as follows: (a) elementary teachers - strong target loading (.66) and negligible nontarget loadings (.01, -.08), (b) secondary teachers - moderate target loading (.39), with a comparable secondary loading (.37) on DP, and adequate loading on PA (.06), and (c) university teachers - very modest target loading (.29) with a stronger loading on DP (.45) and acceptable loading on the PA factor (.01).

Additionally, four other items cross-loaded, differentially so, across teacher groups; these items were: (a) Item 12, designed to measure PA, consistently cross-loaded onto EE for elementary, intermediate, and secondary teachers, (b) Item 20, designed to measure EE, cross-loaded onto DP for intermediate and university teachers, (c) Item 21, designed to measure PA, cross-loaded modestly onto the DP factor for elementary and intermediate teachers, and (d) Item 22, designed to measure DP, cross-loaded onto the PA factor for elementary teachers, but onto the EE factor for intermediate and secondary teachers.

Finally, although findings were consistent in demonstrating a moderate correlation between the factors of EE and DP for intermediate ($r=.55$), secondary ($r=.43$), and university ($r=.43$) teachers, this was not the case for elementary teachers; the relation was shown to be minimal ($r=-.18$). These results suggest that the structure of burnout may be somewhat different for elementary teachers, than for teachers at other levels of the educational system. A summary of statistics related to the

FA analyses is reported in Table 2.

Insert Table 2 about here

Although the EFA findings suggest that the seven selected items might be problematic when used with particular teacher groups, the true test must come from a CFA approach to the analyses. Indeed, EFA is limited in its ability to: (a) yield unique factorial solutions, (b) define a testable model, (c) assess the extent to which an hypothesized model fits the data, and suggest alternative parameterization for model improvement and, (d) adequately test factorial invariance across groups (Fornell, 1983; Long, 1983; Marsh & Hocevar, 1985). CFA, on the other hand, yields this information and is therefore a more powerful test of factorial validity. We turn now to the results of these CFA analyses.

Confirmatory Factor Analyses

The CFA model in the present study hypothesized a priori that: (a) responses to the MBI could be explained by three factors, (b) each item would have a non-zero loading on the burnout factor it was designed to measure, and zero loadings on all other factors, (c) the three factors would be correlated and, (d) the uniqueness terms for the item variables would be uncorrelated. (The term "uniqueness" is used here in the factor analytic sense to mean a composite of specific and random

measurement error which, in cross-sectional studies, cannot be separated; for an extended discussion, see Gerbing & Anderson, 1984). Based on the same item responses as in the EFA, CFAs were conducted separately for elementary, intermediate, secondary, and university teachers using the LISREL VI program (Joreskog & Sorbom, 1985).

Not surprisingly, based on the EFA findings, the fit of the hypothesized 3-factor model (Model 1), for all teacher groups, was poor from both a statistical (χ^2 values) and a practical (TLI, BBI values) perspective; this model was therefore rejected. These results are presented in Table 3. (Model 0 argues that each item represents a factor and therefore represents the 22-factor model that is needed to compute the TLI and BBI).

Insert Table 3 about here

To identify the misfit in the initially postulated model, a sensitivity analysis (Byrne, 1988a; Byrne, Shavelson, & Muthen, 1989) was conducted. This procedure serves two purposes: (a) to guide the researcher in determining the best-fitting model both statistically and substantively, and (b) to allow the researcher to investigate, under alternative specifications, changes in important measurement (factor loading) and structural parameter (factor variance and covariance)

estimates. Thus, model fitting was continued for each teacher group until the hypothsized model could not be rejected (i.e., $p > .05$). As such, a series of alternatively specified nested models were estimated and their fit compared with the initially hypothesized model (Model 1). The difference in χ^2 ($\Delta\chi^2$) is itself χ^2-distributed, with degrees of freedom equal to the difference in degrees of freedom ($_{\Delta}$df) and can therefore be evaluated statistically; a significant $\Delta\chi^2$ indicating a substantial improvement in model fit. Model respecification and reestimation focused on the modification indices (MIs) provided by the LISREL program for each specified model[2]. As such, parameters having the highest MIs and for which it made substantive sense to do so, were cumulatively relaxed, one at a time. (For a more detailed description and discussion of this proceure in general, see Byrne, 1989; Byrne et al., 1989; with specific application to a single measuring instrument, see Byrne, 1988a).

Before elaborating on the results of these post hoc analyses, however, it is important to point out that once Model 1 was rejected, the analyses ceased to be of a confirmatory nature. The search for a well-fitting model represents an exploratory procedure and carries with it, the usual problems related to the capitalization on chance factors and, thus, the risk of Type I and/or Type II errors. While some have severely criticized this procedure (e.g., Browne, 1982; Cliff, 1983;

Fornell, 1983; MacCallum, 1986), others have argued that it can yield a wealth of substantively meaningful information (e.g., (Bentler & Chou, 1987; Byrne et al., 1989; Tanaka & Huba, 1984). Verification of post hoc models, however, must come from cross-validation with an independent sample.

Turning to Table 3 again, let us now examine results related to the final best-fitting model (Model 2) for each teacher group. In arriving at this final model, the specification of several alternative models (elementary = 28; intermediate = 24; secondary = 20; university = 25) were required. Consistent with the EFA findings, Model 2, for each group, comprised some secondary loadings[3], this number being specific to the particular group under study. Additionally, and as expected, correlated errors played a large part in attaining the model of best fit. These parameters most often represent nonrandom measurement error due to method effects associated with the response format of measuring instruments and are therefore highly likely in the CFA of a single measuring instrument. Indeed, previous research with psychological constructs in general (Bentler & Chou, 1987; Joreskog, 1982; Newcomb, Huba, & Bentler, 1986; Tanaka & Huba, 1984), and with measuring instruments in particular (Byrne, 1988b, 1988c; Byrne & Schneider, 1988) has demonstrated the need to allow for correlated errors in order to attain a well-fitting model. In fact, Bentler and Chou (p. 108) argue that the specification of

a model that forces these error parameters to be uncorrelated is rarely appropriate with real data.

At first blush, one might easily conclude that the final model, as indicated by the χ^2, χ^2/df ratio and TLI values, represented a satisfactory fit to the data for each teacher group; this, however, was not so for elementary teachers. Close inspection of the primary measurement and structural parameters identified vast fluctuations in their estimates throughout the model fitting process; this was not the case with the other three groups. Such instability of parameter estimates is further evidence of an ill-fitting model. It is possible, however, that the somewhat smaller sample size, together with a moderately kurtotic sample distribution (mean KU = 3.08) contributed to the problematic model fit for elementary teachers (see Browne, 1982; Muthen & Kaplan, 1985).

Because final best-fitting models can incorporate a number of parameters, many of which often contribute trivially to model improvement (Bentler & Bonett, 1980), they are typically not used in testing for multigroup invariance; rather, baseline models are determined and then used for this purpose (see Byrne, 1988a; Byrne et al., 1989). A baseline model is the most parsimonious, albeit substantively meaningful model to plausibly represent the data. The establishment of a baseline model, specific to each group under study, is an important prerequisite to testing for invariance across groups. This

model (Model 3) is summarized in Table 3 for intermediate, secondary, and university teachers. Due to the problematic fit of the originally hypothesized model for elementary teachers, 23 model respecifications were required in order to attain a TLI of .90; this model included four secondary loadings and, thus, model complexity as well as non-parsimony became an important issue. Given these findings, together with the sample distribution problems noted earlier, the determination of a baseline model for this group would have been ill-founded and, therefore, was not considered.

Several criteria guided the selection of the baseline models presented in Table 3. First, each model exhibited a reasonably good fit to the data, as indicated by a TLI >.90; the TLI was considered preferable to the BBI as a threshold since the latter has been shown to be sample size dependent (Marsh et al., 1988). Second, all major parameter estimates (factor loadings, variances and covariances, and error variances) were reasonable and statistically significant (see Byrne, 1989; Joreskog & Sorbom, 1985). Third, all secondary loadings and error covariances (i.e., correlated errors) were substantial (>.20) and significant (see Byrne, 1988a; Byrne et al., 1989); this was not so for subsequently specified models. Fourth, the measurement and structural parameter estimates remained stable despite the addition of the secondary loadings and correlated errors (see Bagozzi, 1983; Fornell, 1983;

MacCallum, 1986). Finally, factor loadings and factor variance/covariances in the baseline model correlated >.94 and .99, respectively, with the final best-fitting model, thereby substantiating the stability of parameter estimates in the baseline model (see Bentler & Chou, 1987; Byrne et al., 1989; Newcombe et al., 1986); these correlation values, for intermediate, secondary, and university teachers, are footnoted in Table 3.

In examining Table 3, we see that the baseline models differ for each group of teachers with respect to both the number and specification of secondary loadings and correlated errors. This is because instruments are often group-specific in the way they operate and, thus, these models are not expected to be identical across groups. As expected, correlated errors played a prominant role in the establishment of baseline models; typically, their addition resulting in a substantially significant drop in χ^2. For example, the specificaton of a correlated error term between Items 1 and 2 measuring EE, for intermediate teachers, resulted in a $\Delta\chi^2 = 74.88$; a second specified correlated error between Item 15 measuring DP, and Item 16 measuring EE, resulted in a drop of 53.27 in the value. These parameters, as noted earlier, represent a method effect associated with the response format of the MBI.

Of more importance, from a substantive standpoint, were the secondary loadings needed to obtain a well-fitting baseline

model. In general, these parameter specifications were consistent with findings from the EFA and, thus, substantiate problems associated with particular items; again, problematic items were group-specific. For both intermediate and secondary teachers, Item 12 (designed to measure PA) cross-loaded onto the EE factor. For secondary teachers, an additional item (Item 14) measuring EE, also loaded onto PA. Interestingly, four different secondary loadings defined the baseline model for university professors; these were: (a) Item 20 (target: EE) cross-loaded onto both the DP and PA factors, (b) Item 11 (target: DP) cross-loaded onto EE, and (c) Item 16 (target: EE) cross-loaded onto DP.

Once the baseline models were established, analyses proceeded in testing for the equivalency of the MBI across teacher groups. Given the more complex baseline structure for university teachers compared with intermediate and secondary teachers, however, tests of invariance were limited to the latter two groups only.

Tests of Invariance

Of primary interest in testing for the invariance of a measuring instrument is the equivalence across groups of the item scaling units and latent factor covariances. Noninvariant scaling units are an indication that the items are differentially valid across groups; noninvariant factor covariances suggest that the structure of the construct

underlying the measurement instrument varies across groups.

Tests of invariance involved specifying a model in which certain parameters were constrained to be equal across groups and then comparing that model with a less restrictive model in which these parameters were free to take on any value. As with model-fitting, the $\Delta\chi^2$ between competing models provided a basis for determining the tenability of the hypothesized equality constraints; a significant $\Delta\chi^2$ indicating noninvariance. Space limitations preclude further elaboration of these invariance testing procedures. However, detailed elsewhere, are descriptions of the procedure in general (Byrne, 1989; Byrne et al., 1989; Marsh & Hocevar, 1985), and bearing on single measuring instruments in particular (Byrne, 1988a, 1988b, 1988c).

The simultaneous testing for invariance across intermediate and secondary teachers revealed the equivalence of all items, including the one secondary loading common to both groups (the cross-loading of Item 12 onto the EE factor)[5]; the other secondary loading for secondary teachers was, of course, not constrained equal across groups. Interestingly, tests for invariant factor covariances yielded significant differences between intermediate and secondary teachers in relations between EE and DP, and between DP and PA; the EE/PA relation was group-invariant. These findings point to a differential structure of burnout for intermediate and secondary teachers.

Whereas DP was found to be highly correlated with EE for intermediate teachers, it was only moderately so for secondary teachers; conversely, while relations between EE and PA, and between DP and PA were low for intermediate teachers, they were moderately strong for secondary teachers. Results from the series of invariance tests are summarized in Table 4, and latent factor correlations are presented in Table 5.

Insert Tables 4 and 5 about here

Discussion

The present study investigated the factorial validity of the MBI, separately, for teachers at four levels of the educative system --- elementary, intermediate, high school, university, and then tested for the equivalence of item measurements and inter-factor relations across intermediate and secondary teachers. EFA was used to explore the initial factor structure, and CFA to validate the hypothesized structure and and test for its invariance across groups; these procedures represent three incrementally powerful tests of factorial validity.

In general, although the EFA procedures yielded findings that were less than optimal for intermediate and secondary teachers, the more rigorous and powerful CFA analyses demonstrated a remarkably well-defined 3-factor model for both.

Nonetheless, one item appears to require some attention. For both groups of teachers, Item 12 measuring PA, loaded onto the EE factor; EFA and CFA results were unanimous with respect to this item. Indeed, from the perspective of content validity, one could easily construe this item as a measure of the latter construct. For secondary teachers, the CFA loading of Item 14 (designed to measure EE) onto PA is more difficult to explain, especially since the EFA results yielded the expected loading. At present, I can only explain this anomaly as an artifact of the present sample; I suspect this finding will disappear upon replication.

The factorial validity of the MBI for university professors, although somewhat less psychometrically glowing than for intermediate and secondary teachers, may still be considered reasonably sound. Problematic here, however, were four substantially cross-loaded items, all of which were identified in both the EFA and CFA analyses. The fact that Item 20 (EE) loaded on both the DP and PA factors, serves notice that the item is a weak measure at best, of its target factor. Indeed, its cross-loading onto DP accounted for the largest drop in X^2 among all the post hoc respecifications ($\chi^2(1) = 38.01$). With some justification, there seems also to be a need to examine Item 16, measuring EE. Although the cross-loading of this item onto DP was most dramatic for university professors, it also occurred at a post-baseline stage for

intermediate teachers. Finally, the loading of Item 11 (DP) onto EE appears to be unique to the university population. Although an explanation is hard to come by, this cross-loading accounted for a substantial reduction in χ^2 ($\chi^2(1) = 20.23$) and, therefore bears further investigation.

For the present sample of elementary teachers, the factorial structure of the MBI was somewhat less than optimal. It appears that Items 10 and 11, designed to measure DP, require some modification; each loaded negligibly on this factor in both the EFA and CFA analyses. Aside from these poor loadings, however, other serious cross- loadings are believed to have contributed to the poor fit of the MBI to the present data. Further analyses using procedures capable of yielding elliptical estimators appropriate for the analysis of kurtotic data are needed before drawing conclusions on use of the MBI with elementary teachers.

The fact that tests for invariance revealed all item measurements, including the secondary loading of Item 12, across intermediate and secondary teachers, is quite remarkable! This indicates that, aside from the differential method effects and additional secondary factor loading for high school teachers, all items were perceived identically by both groups of educators. Indeed, these results speak highly for the psychometric soundness of the MBI in measuring burnout for teachers at both the intermediate and secondary levels.

The revelation of a differential burnout structure for these two groups of teachers, albeit unexpected, is rather interesting (see Table 5). Although the directional pattern of correlations among EE, DP, and PA was the same for both, the strength of relations differed, and significantly so for relations between EE and DP, and between DP and PA. The results, for secondary teachers, were very similar to those reported by Maslach and Jackson (1986) based on a combination of occupations. As such, feelings of emotional exhaustion were associated with a moderately high and positive degree of depersonalization, whereas the relation with feelings of personal accomplishment was negative and moderate; feelings of depersonalization were moderately and negatively related to feelings of personal accomplishment.

An important finding of this study has been evidence that intermediate and secondary teachers appear to be the most closely matched with respect to validity bearing on the MBI. In light of the rigorous statistical approach to the present analyses in general, and the stringency of LISREL CFA procedures in particular, the MBI has demonstrated sound measurement properties for use with these two levels of teachers. It appears clear, however, that elementary teachers and university professors march to a different drummer; further validity research is needed with both of these populations, and is currently under way by the present author.

References

Alwin, D.F. & Jackson, D.G. (1980). Measurement models for response errors in surveys: Issues and applications. In K.F. Schuessler (Ed.), Sociological Methodology (pp. 68-119). San Francisco: Jossey-Bass.

Bagozzi, R.P. (1983). Issues in the application of covariance structure analysis: A further comment. Journal of Consumer Resources, 9, 449-450.

Beck, C.L. & Gargiulo, R.M. (1983). Burnout in teachers of retarded and nonretarded children. Journal of Educational Research, 76, 169-173.

Belcastro, P.A., Gold, R.S., & Hays, L.C. (1983). Maslach Burnout Inventory: Factor structures for samples of teachers. Psychological Reports, 53, 364-366.

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P.M. & Chou, C.P. (1987). Practical issues in structural modeling. Sociological Methods & Research, 16, 78-117.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), Topics in applied multivariate analysis (pp. 72-141). New York: Cambridge University Press.

Byrne, B.M. (1988a). Testing the factorial validity and invariance of a measuring instrument using LISREL confirmatory factor analyses: A reexamination and application. Multiple Linear Regression Viewpoints, 16, 33-80.

Byrne, B.M. (1988b). Measuring adolescent self-concept: Factorial validity and equivalency of the SDQ III across gender. Multivariate Behavioral Research, 24, 361-375.

Byrne, B.M. (1988c). The Self Description Questionnaire III: Testing for equivalent factorial validity across ability. Educational and Psychological Measurement, 48, 397-406.

Byrne, B.M. (1989). A Primer of LISREL: Basic applications and programming for confirmatory factor analytic. New York: Springer-Verlag.

Byrne, B.M. & Schneider, B.H. (1988). Perceived Competence Scale for Children: Testing for factorial validity and invariance across age and ability. Applied Measurement in Education, 1, 171-187.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456-466.

Caplan, R.D. & Jones, K.W. (1975). Effects of work load, role ambiguity, and type A personality on anxiety, depression, and heart rate. Journal of Applied Psychology, 60, 713-719.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. Multivariate Behavioral Research, 18, 115-126.

Farber, B.A. & Miller, J. (1981). Teacher burnout: A psychoeducational perspective. Teachers College Record, 83, 235-243.

Firth, H., McIntee, J., McKeown, & Britton, P.G. (1985). Maslach Burnout Inventory: Factor structure and norms for British nursing staff. Psychological Reports, 57, 147-150.

Fornell, C. (1983). Issues in the application of covariance structure analysis: A comment. Journal of Consumer Research, 9, 443-448.

Freudenberger, H.J. (1974). Staff burn-out. Journal of Social Issues, 30, 159-165.

Gerbing, D.W. & Anderson, J.C. (1984). On the meaning of within-factor correlated measurement errors. Journal of Consumer Research, 11, 572-580.

Gold, Y. (1984). The factorial validity of the Maslach Burnout Inventory in a sample of California elementary and junior high school classroom teachers. Educational and Psychological Measurement, 44, 1009-1016.

Ianni, F.A. & Reuss-Ianni, E. (1983). "Take this job and shove it!": A comparison of organizational stress and burnout among teachers and police. In B.A. Farber (Ed.), Stress and burnout in the human service profession (pp. 82-96). New York: Pergamon

Iwanicki, E.F. & Schwab, R.L. (1981). A cross validation study of the Maslach Burnout Inventory. Educational and Psychological Measurement, 41, 1167-1174.

Jackson, S.E., Schwab, R.L., & Schuler, R.S. (1986). Toward an understanding of the burnout phenomenon. Journal of Applied Psychology, 71, 630-640.

Joreskog, K.G. (1982). Analysis of covariance structures. In C. Fornell (Ed.), A second generation of multivariate analysis Vol 1: Methods (pp. 200-242). New York: Praeger.

Joreskog, K.G. & Sorbom, D. (1985). LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Mooresville, Ind: Scientific Software.

Long, J.S. (1983). Confirmatory factor analysis. Beverly Hills, CA: Sage.

MacCallum, R. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Marsh, H.W. & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. Psychological Bulletin, 97 (3), 562-582.

Maslach, C. (1982). Burnout: The cost of caring. Englewood Cliffs, NJ: Prentice-Hall.

Maslach, C. & Jackson, S.E. (1981a). Maslach Burnout Inventory Manual. Palo Alto, CA: Consulting Psychologists Press.

Maslach, C. & Jackson, S.E. (1981b). The measurement of experienced burnout. Journal of Occupational Behavior, 2, 99-113.

Maslach, C. & Jackson, S.E. (1984). Burnout in organizational settings. In S. Oskamp (Ed.) Applied social psychology in organizational settings (pp. 133-153). Beverly Hills, CA: Sage.

Maslach, C. & Jackson, S.E. (1986). Maslach Burnout Inventory Manual (2nd edition). Palo Alto, CA: Consulting Psychologists Press.

Muthén, B., & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Newcomb, M.D., Huba, G.J., & Bentler, P.M. (1986). Determinants of sexual and dating behaviors among adolescents. Journal of Personality and Social Psychology, 50, 428-438.

Perlman, B. & Hartman, E.A. (1982). Burnout: Summary and future research. Human Relations, 35, 283-305.

Schwab, R.L. & Iwanicki, E.F. (1982a). Who are our burned out teachers? Educational Research Quarterly, 7, 5-16.

Schwab, R.L. & Iwanicki, E.F. (1982b). Perceived role conflict, role ambiguity, and teacher burnout. Educational Administration Quarterly, 18, 60-74.

Selye, H. (1956). The stress of life. New York: McGraw-Hill.

Tanaka, J.S. & Huba, G.J. (1984). Confirmatory hierarchical factor analysis of psychological distress measures. Journal of Personality and Social Psychology, 46, 621-635.

Tosi, H. & Tosi, D. (1970). Some correlates of role conflict and role ambiguity among public school teachers. Journal of Human Relations, 18, 1068-1079.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Footnotes

1. Selection of the BBI and TLI was based both on their widespread application and on their usefulness in comparing samples of unequal size (Marsh, Balla, & McDonald, 1988).
2. An MI may be computed for each constrained parameter and indicates the expected decrease in X^2 if the parameter were to be relaxed; the decrease, however, may actually be higher.
3. Secondary loadings are synonymous with cross-loadings; they represent measurement loadings onto more than one factor.
4. The somewhat lower BBI values may be an artifact of the sample size since they have been shown to be less independent than the TLI in this regard (Marsh et al., 1988).
5. Item 12 is "I feel very energetic".

Table 1

Target Loadings[1] from Exploratory Factor Analyses

Item Number	Elementary (n=98) Target Loading	Intermediate (n=163) Target Loading	Secondary (n=162) Target Loading	University (n=219) Target Loading
		Emotional Exhaustion		
1	.74	.87	.68	.79
2	.68	.91	.73	.75
3	.69	.67	.72	.80
6	.64	.25	.63	.42
8	.79	.68	.78	.89
13	.67	.58	.60	.68
14	.71	.66	.58	.58
16	.66	.12	.39	.29
20	.78	.45	.71	.54
% Variance explained	28.40	34.90	30.70	28.20
		Depersonalization		
5	.68	.62	.47	.65
10	.17	.67	.70	.44
11	.19	.65	.79	.43
15	.70	.91	.31	.74
22	.52	.36	.33	.68
% Variance explained	9.60	7.40	7.40	8.30
		Personal Accomplishment		
4	.28	.26	.38	.55
7	.56	.52	.66	.72
9	.54	.71	.70	.60
12	.46	.45	.60	.58
17	.53	.66	.69	.72
18	.47	.67	.73	.59
19	.52	.67	.70	.68
21	.33	.37	.51	.54
% Variance explained	11.80	14.20	14.50	16.70

[1] Target loadings are factor loadings on the factor which the item was designed to measure.

Table 2

Summary Statistics for Exploratory Factor Analysis

	Elementary (n=98)	Intermediate (n=163)	Secondary (n=162)	University (n=219)
Target Loadings[1] (22 factor loadings)				
High	.79	.91	.79	.89
Low	.17	.12	.31	.29
Median	.50	.66	.69	.63
% ≥ .40	82%	82%	82%	95%
Non-target Loadings (66 factor loadings)				
High	.64	.74	.37	.45
Low	.00	.00	.00	.00
Median	.04	.05	.05	.01
% ≥ .40	5%	3%	0%	2%
Factor Correlations (6 factor correlations)				
High	-.18	.54	.43	.43
Low	-.02	.01	-.20	.08
Median	-.16	-.08	-.32	.23
% ≥ .40	0%	33%	33%	33%

[1] Target loadings are factor loadings on the factor which the item was designed to measure; nontarget loadings refer to all other factor loadings.

Table 3

Summary of Major Confirmatory Factor Analytic Models[a]

	Competing Models	χ^2	df	$\Delta\chi^2$[b]	Δdf	χ^2/df	TLI	BBI
				Elementary				
0	Null Model	1086.71	231	—	—	4.68	—	—
1	Hypothesized 3-factor model	505.51	206	—	—	2.45	.61	.54
2	Final Model[c]; Model 1 that also included 6 secondary loadings and 21 correlated errors	211.23	179	294.28***	27	1.18	.95	.81
3	Baseline Model[d]	—	—	—	—	—	—	—
				Intermediate				
0	Null Model	2178.88	231	—	—	9.43	—	—
1	Hypothesized 3-factor model	638.52	206	—	—	3.10	.75	.71
2	Final Model[c]; Model 1 that also included 6 secondary loadings and 17 correlated errors	214.36	183	424.16***	23	1.17	.98	.90
3	Baseline Model[e]; Model 1 that also included 1 secondary loading and 10 correlated errors	327.28	195	311.24***	11	1.68	.92	.85
				Secondary				
0	Null Model	1687.01	231	—	—	7.30	—	—
1	Hypothesized 3-factor model	461.93	206	—	—	2.24	.80	.73
2	Final Model[c]; Model 1 that also included 3 secondary loadings and 16 correlated errors	218.73	187	243.20***	19	1.17	.97	.87
3	Baseline Model[f]; Model 1 that also included 2 secondary loadings and 13 correlated errors	243.78	191	218.15***	15	1.28	.96	.86

Table 3 Cont'd

Competing Models	χ^2	df	$\Delta\chi^2$ [b]	Δdf	χ^2/df	TLI	BBI
			University				
0 Null Model	2212.80	231	—	—	9.57	—	—
1 Hypothesized 3-factor model	548.74	206	—	—	2.66	.81	.75
2 Final Model[c]; Model 1 that also included 5 secondary loadings and 19 correlated errors	212.81	182	335.93***	24	1.17	.98	.90
3 Baseline Model[g]; Model 1 that also included 4 secondary loadings and 4 correlated errors	357.58	198	191.15***	8	1.81	.91	.84

*** $p < .001$

[a] For sake of simplicity and clarity, only 2 of a series of post hoc models are presented here, in addition to the null and initially hypothesized models

[b] Represents difference from Model 1

[c] Probability ≥ 0.05

[d] Given the ill-fitting final model, no baseline model could be established

[e] Factor loadings correlated .97 and factor variances/covariances correlated .99 with final model

[f] Factor loadings correlated .95 and factor variances/covariances correlated .99 with final model

[g] Factor loadings correlated .93 and factor variances/covariances correlated .99 with final model

Table 4

Tests for Invariance Across Intermediate and Secondary Teachers

	Competing Models	χ^2	df	$\Delta\chi^2$	Δdf	p
1	Number of factors invariant	571.06	386	—	—	NS
2	Items measuring EE invariant	582.09	394	11.03	8	NS
3	Model 2 with secondary loading of Item 12 loading onto EE invariant	584.70	395	13.64	9	NS
4	Model 3 with items measuring DP invariant	590.36	399	19.30	13	NS
5	Model 4 with items measuring PA invariant	599.57	406	28.51	20	NS
6	Model 5 with all variances/ covariances invariant	622.10	412	51.04	26	<.01
7	Model 5 with all variances invariant	610.32	409	39.26	23	<.05
8	Model 5 with all covariances invariant	620.46	409	49.40	23	<.001
9	Model 5 with EE/DP covariance invariant	608.92	407	37.86	21	<.05
10	Model 5 with EE/PA covariance invariant	602.50	407	31.44	21	NS
11	Model 5 with DP/PA covariance invariant	604.70	407	33.64	21	<.05

EE = Emotional Exhaustion; DP = Depersonalization; PA = Personal Accomplishment

Table 5

Baseline Latent Factor Correlations for Intermediate and Secondary Teachers[a]

Factors of Burnout	Emotional Exhaustion	Depersonalization	Personal Accomplishment
		Intermediate	
Emotional Exhaustion	1.00		
Depersonalization	.81	1.00	
Personal Accomplishment	-.14	-.12	1.00
		Secondary	
Emotional Exhaustion	1.00		
Depersonalization	.64	1.00	
Personal Accomplishment	-.36	-.47	1.00

[a] Standardized Solution